**Overall Information**

* We will complete this assignment **together** in class over the course of two weeks. You should submit the final assignment here.
* You will be scored as complete/incomplete for this assignment.
* You must submit a report with code and interpretation within the same document. This report should be created within markdown or jupyter notebooks.

**Find Text**

* As a class, we will find a text source to analyze to predict a categorical outcome.
* Import the text into your report.

**Set up Text**

* Do preprocessing on your text to prepare it for the final machine learning model.
* What items do you think will be important in your preprocessing to clean?

**Create Feature Extractions**

* Create a “one-hot” encoding using the count vectorizer and binary options.
* Create the bag of words encoding using the count vectorizer.
* Create the TF-IDF normalization using the tfidf vectorizer.
* Create four word2vec models:
  + Using 50 dimensions versus 100 dimensions.
  + Using cbow and skipgram embeddings.
  + Using a 5 window size.
* Extract the spacy information for your documents.